

# The Strength of Inferences About Causes of Trends in Bird Populations

Frances C. James  
Department of Biological Science  
Florida State University  
Tallahassee, FL 32306-2043, USA

Charles E. McCulloch  
Biometrics Unit  
337 Warren Hall  
Cornell University  
Ithaca, NY 14853, USA

BU-1233-M

April 1994

The Strength of Inferences About Causes of Trends in Bird Populations

Frances C. James<sup>1</sup>

and

Charles E. McCulloch<sup>2</sup>

For T. E. Martin and D. M. Finch (editors) Ecology and management of  
Neotropical migratory birds: a synthesis and review of the  
critical issues. Oxford University Press. *Submitted March 1984*

<sup>1</sup>Department of Biological Science, Florida State University,  
Tallahassee, FL 32306-2043, USA

<sup>2</sup>Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, NY 14853,  
USA

## ABSTRACT

Our objective in this paper is to explore ways in which the logic of experimental design might be applied to the analysis of causes of trends in bird populations. The general problem is how to set up comparisons of observations that improve the power of causal inference in situations that do not permit true controlled experiments (those with randomized assignment of experimental units to treatments, control groups, and replication). It is important to recognize and acknowledge that, generally, only weak inferences will be possible, but even weak inferences are preferable to simple conjecture. We illustrate several quasi-experimental designs (those without random assignment) and discuss how they differ in the level of causal inference allowed. Principles of quasi-experimental design can be used both in field experiments and in purely observational studies. Depending on the types of comparisons made, the effects of some of the potentially competing explanations can be removed.

With only the results of a field experiment, which is usually conducted in an area that is small relative to the area of interest, the researcher would not know how general the result might be, and from only the results of a comparative observational study, the researcher would be left with weak inferences. But a combination of field experiments and comparative observational studies can provide both a reasonable strength of inference of causality and a reasonably general scope of inference. At both the level of the local experiment and that of the observational study, we emphasize the importance of making comparisons

that allow for the control of extraneous causes by the incorporation of principles of experimental design to whatever extent possible.

#### INTRODUCTION

The idea that land birds that breed in the United States and Canada and spend the winter in the New World tropics have been suffering severe population declines is the basis for the United States government-sponsored program Partners in Flight and is the rationale for the publication of this volume. We believe that the evidence of an overall decline in populations of Neotropical migrants is weak (James et al. ms). Analyses of the best data for population trends in North American land birds for the past 25 years, the Breeding Bird Survey, indicate general declines in some species of Neotropical migrants and increases in others, but no overall pattern that pertains to the group as a whole (Peterjohn et al. this volume). However, documentation of population trends is not the issue in this paper. Instead we will address methods for the confirmation of hypothesized causes of population trends, once population trends have been established. The confirmation of causes is more a matter of logic than of data analysis. It involves the design of the study, how comparisons are made among observations, rather than the significance of statistical tests. Of course, both steps, documentation of trends and determination of their causes, should precede decisions about management.

Controlled experiments are usually considered to be the only way to determine cause-and-effect relationships, but true field experiments (including random assignment of experimental units to treatments) are often not possible, and one must be satisfied with either

quasi-experiments or with observational studies. In this sense, most field experiments are, in fact, quasi-experiments. (For definitions of terms, see Table 1, and for notation that makes the logic of causal reasoning explicit, see Table 2.) In addition, the external validity of field experiments, that is, the extrapolation of the results to domains larger than that of the experiment, is a matter of judgment rather than rigorous statistical inference (Eberhardt and Thomas 1991). What we recommend is a further step, that principles of quasi-experimental design be applied to comparative observational studies. Although such efforts will provide even weaker causal inferences than would quasi-experiments, they have the important advantage that they can incorporate the entire domain of interest. Just as with the new nonmanipulative comparative methods in evolutionary biology (Harvey and Pagel 1991), studies of population regulation could benefit from advances in the development of comparative methods. Carefully designed comparative observational studies (the analytical sampling of Eberhardt and Thomas 1991) should not be viewed as second-class experiments. They have their own justification (Kish 1987).

Referring to the field of epidemiology, another area that seeks to find the causes of broad-scale phenomena, Rothman (1988) wrote that, without such effort toward falsification, we risk forming a consensus based on shared irrational belief buoyed by an accumulation of supporting observations. Maclure (1988) added that, without attention to alternative explanations, we are left with untested beliefs and uncritical convictions about what should be done. Our argument here is that any real confirmation of a cause of a population trend would have

to be made on the basis of a combination of the results of controlled field experiments, which would provide inference about a local situation, and the results of comparative observational studies, which would provide information about processes in the large intact ecosystem.

#### RESEARCH DESIGNS FOR CAUSAL ANALYSIS

Our definitions (Table 1) and notation (Table 2) are based primarily on those of Campbell and Stanley (1966), Cook and Campbell (1979), and Cochran (1983), and we cite alternative terms used by other authors. Campbell and Stanley (1966) used the following notation:

X = exposure of a group to a treatment or an event

O = an observation

R = random assignment of experimental units to treatments.

Rows of X's and O's are for single groups, and columns are for simultaneous treatments and observations. Replication (multiple experimental units per group) would ordinarily be included and is not discussed here. It affects the likely statistical significance of effects but not the strength of causal inference that can be made once effects have been detected. The first three categories in Table 2 are experimental in that the researcher applies the treatment. The fourth category, comparative observational studies, involves making planned comparisons of sets of observations of naturally occurring events. The random assignment of experimental units to treatments is only possible with category 2, an example of a true experimental design. We estimate the strength of inference possible with each design on a subjectively determined scale from 0 to 5.

### Inadequate Designs

We define controlled experiments as those cases in which the researcher is in control of the application of treatment. The first three experimental designs in Table 2, the one-shot case study,

X O,

the one group pretest-posttest,

O X O,

and the static group comparison,

X O

O,

are inadequate for causal inference, even if the work is based on controlled experiments, because they offer insufficient control over extraneous processes. If the researcher wished to conclude that the observation after the treatment was affected by the treatment, that conclusion would have to be confirmed from information beyond the experiment, because the logic of the design was inadequate for causal inference. In the first case, one group was studied once, so there was no control and no comparison. In the second case, one group was studied before and after a treatment. This design controls for initial inequalities by using the group as its own control group, but does not control for alternate causes that occurred at the same time as X. In neither case is there any assurance that only the treatment caused the effect. In the third case, there is a control group that did not receive the treatment, but because there was no random assignment, there is no assurance that the initial groups did not differ before the treatment.

### A True Experimental Design

We give only one example of a true experimental design (Table 2, category #2):

R O X O

R O O,

the pretest-posttest control-group design. Here the R's signify that, before the experiment, experimental units were randomly assigned to treatments. A comparison of the differences between observations before and after the treatment and differences without the treatment is a test for the effect of the treatment.

### Quasi-experimental Designs

The term quasi-experiment refers to an experiment in which full experimental control is lacking because the random assignment of experimental units to treatments was not performed (Table 1). Inferences from such experiments are necessarily weaker than would be those from a true experiment. According to Cook and Campbell (1979, p. 6) the basic issue with quasi-experiments is that

"comparisons depend on non-equivalent groups that differ from each other in many ways other than the presence of a treatment whose effects are being tested. The task confronting persons who try to interpret the results from quasi-experiments is basically one of separating the effects of a treatment from those due to initial noncomparability between average units in each treatment group; only the effects of the treatment are of research interest. To achieve this separation of effects, the researcher has to explicate the specific threats to valid causal inference that a random



assignment rules out and then in some way deal with these threats.

In a sense, quasi-experiments require making explicit the irrelevant causal forces hidden within the ceteris paribus of random assignment."

With the nonequivalent-control-group design, in which both groups are observed pretest and posttest,

O X O

O O,

the only difference from the example of a true experimental design is that there is no preexperimental randomization to assure sampling equivalence. Because of the simultaneous observations made of the two groups, the design removes the effect of time. But, because of the lack of randomization, even if a statistical comparison, similar to the one for the true experiment, is significant, it does not necessarily test the effect of the treatment (see Hurlbert 1984).

A single time series of observations into which the observer has introduced an intervention,

O O O O X O O O O

is another type of quasi-experimental design. In this case there is no control group, but the multiple observations pretest and posttest help the researcher judge the effect of the treatment. The weakness of this design is that it does not remove the effects of competing causes occurring at the same time as X.

With the inclusion of a control group that is not subjected to the intervention, the single-time-series design becomes the multiple-time-series design,

0 0 0 0 X 0 0 0 0  
 0 0 0 0 0 0 0 0,

and substantial strength of inference is gained. If a researcher is unable to perform a true experiment, the multiple-time-series design is a particularly strong alternative.

It is possible to extend all of these designs to cover tests of alternative causes. For example with the multiple-time-series design and treatments X and Y, the notation would be

0 0 0 0 X 0 0 0 0  
 0 0 0 0 Y 0 0 0 0  
 0 0 0 0 0 0 0 0  
 0 0 0 0 XY 0 0 0 0.

The notation XY means that both treatments X and Y occurred together.

#### Comparative Observational Studies

Observational studies are not experimental because the researcher is not in control of treatments. Even so, naturally occurring events can be viewed as treatments. In some cases principles of quasi-experimental design, because they help the researcher gain some control over irrelevant causal forces, can be used to strengthen causal inferences (Cook and Campbell 1979; Table 2, category #4). Of course inadequate experimental designs (Table 2, category #1) will be inadequate in these cases as well. Many causal hypotheses about naturally occurring events are first addressed through inadequate designs, so the question becomes how to move them forward through a progression of designs that provide stronger causal inference. The category of comparative observational studies addresses this issue, how

some rival hypotheses about causes can be rendered implausible by the design of the comparisons (Cook and Campbell 1979). This category includes the "natural experiments" of Diamond (1986).

When there have been no abrupt natural events, the analysis of causes (treatments) in comparative observational studies is the most difficult of all, yet this is the situation in many cases of the analysis of trends in bird populations. Even if multiple observations are available, the analysis may require quantification of static levels of hypothesized causes. For two hypothesized causes, the design is likely to be

X	0	0	0	0
Y	0	0	0	0
XY	0	0	0	0
	0	0	0	0

which would have to be interpreted as a static group comparison with multiple observations, a weak design.

Suppose we wish to separate the effects of changing levels of cowbird parasitism (X) and habitat fragmentation (Y) on population trends of the Wood Thrush (Hylocichla mustelina). We would need measurements of the degree of parasitism and fragmentation for each of a number of sites. Our actual design would be complicated because parasitism and fragmentation would probably be continuous measures rather than just absent (control) or present (X or Y). So with population size measurements (O) in consecutive years (e.g., as in BBS data) and continuous (measured) levels of parasitism ( $X_1, X_2, \dots, X_n$ ) and fragmentation ( $Y_1, Y_2, \dots, Y_n$ ) the design would be:

$$0 \ X_1 Y_1 \ 0$$

$$0 \ X_2 Y_2 \ 0$$

$$0 \ X_3 Y_3 \ 0$$

$$\dots$$

$$\dots$$

$$\dots$$

$$0 \ X_n Y_n \ 0$$

With treatments that were merely present or absent and observations before and after treatments, it would have been easy to calculate the effect of parasitism (by comparing the change in the XY group to the change in the Y group and/or X to control) or fragmentation (by comparing XY to X and/or Y to control), nonequivalent-control-group designs. However, because the levels of each factor vary among groups, a multiple-regression model must be used to try to infer the effect of each factor independently of the other. Unfortunately, this is a common situation, and the validity of the inferences "will depend not only on the study design but also on the skills and ability of the investigator to develop sound subjective interpretations of the test results" (Skalski and Robson 1992). Incorrect assumptions incorporated in the regression model (which is necessarily a simplification of the true situation) may result in biased estimates of causes.

#### Correlation and Causation

The arguments in this paper are relevant to the commonly stated idea that one cannot infer causation from correlations among variables or, more properly, that correlational analysis alone is an insufficient basis for inferring causation. With observational studies, the problem

is that it may be difficult to control possibly confounding factors. The way to improve such control is less a matter of the method of analysis (e.g., correlation, regression) than of the design of the study (e.g., single time series, multiple time series). Correlational analyses based on a strong design can provide causal inference. For discussion of appropriate data analysis to accompany various experimental designs, see Scheiner and Gurevitch (1993).

#### EXAMPLES OF QUASI-EXPERIMENTS AND COMPARATIVE OBSERVATIONAL STUDIES WITH COMMENTS ON THE STRENGTH OF THE CAUSAL INFERENCES THAT THEY ALLOW

In this section we present four examples of studies that attempt to identify causes of population trends in birds, and we show how their designs fit into the categories in Table 2. The first two examples are local quasi-experiments, the third and fourth examples are broad-scale comparative observational studies. The example of the Kirtland's Warbler (Dendroica kirtlandii) uses a single-time-series design with two interventions due to two different types of treatments. The example of the Great Tit (Parus major) demonstrates the added strength of inference that might have been gained with a multiple-time series design.

With comparative observational studies, much larger geographic areas can be considered. The third example, a comparative observational study of population trends in the Ovenbird (Seiurus aurocapillus), uses Breeding Bird Survey (BBS) data. It begins with an inadequate design but is somewhat strengthened by an independent analysis. The fourth example, the Carolina Wren (Thryothorus ludovicianus), uses BBS data in a comparative observational study with a multiple-time-series design. Results for different physiographic strata are viewed as groups, and the

effects of a naturally occurring perturbation, severe winter weather, had different levels of impact in different strata.

1. A Quasi-experimental Design, Single Time Series: Kirtland's Warbler

The endangered Kirtland's Warbler has only one known breeding population, the one in early successional stands of jack pines (Pinus banksiana) mainly in two counties in central Michigan. Field surveys showed that this population declined by 50% in the 1960's and that brood parasitism by the Brown-headed Cowbird (Molothrus ater) was severe. With the establishment of a sustained intensive cowbird control program in the 1970's, the number of singing male warblers stabilized at about 200 individuals, but no major increases in the breeding population occurred (Probst 1986). In 1980 a prescribed fire got out of hand and burned several thousand acres, setting into motion an extensive natural regeneration of jack pines. By the late 1980's the young forest in the Mack Lake Burn had become suitable habitat for the Kirtland's Warbler, and since that time the breeding population has increased dramatically (Fig. 1). The continuing cowbird-control program is probably necessary, and the conservation of the species will probably be well served by a program of habitat manipulation that involves exceptionally hot fire and natural regeneration of jack-pine forest. This single time series, a quasi-experimental design with no control and two interventions,

0 0 0 0 X 0 0 0 0 Y 0 0 0 0 0,

allows moderately strong inferences even though it does not remove the effects of a competing cause occurring at the same time as X or Y. We rank its strength of inference a 3 (Table 2, 3b). Had there been a

comparison group that had not been subjected to either cowbird control or the hot fire, stronger inference would have been possible.

2. A Quasi-experimental Design, Single Time Series: Great Tit

Tinbergen et al. (1985) provide an example of the danger of misinterpreting the results of field experiments that offer only moderate strength of inference and that do not consider alternative causes. The data are from the classic long-term studies of fluctuations in a population of Great Tits using nest boxes on the island of Vlieland in The Netherlands. Kluyver and coworkers had experimentally reduced the numbers of fledglings in the breeding seasons of 1960-1963 and 1967-68 and found that adult survival was increased in the years following manipulations. This result was interpreted by Kluyver (1971, as cited by Tinbergen et al. 1985) as evidence of density-dependent population regulation. Tinbergen et al. (1985) reanalyzed the data, adding information up to 1978 that included further manipulations from 1970 to 1974 (Fig. 2). They also used data for annual variation in the seed crop available to birds for food in winter. The reanalysis showed that years when there were manipulations also happened to be years with abundant winter food. Adult survival was, in fact, positively related to the seed stock in the previous winter and the number of fledglings produced per pair in the previous breeding season. Thus, winter food supply was as valid an explanation of variation in adult survival as was density-dependent population regulation. Had there been a control population that was not manipulated, and thus a multiple-time-series rather than a single-time-series design, the presence or lack of causal association between the experimental conditions and adult survival would

have been clear from the beginning. This example shows two things, (1) the danger, even with experiments, of not considering possibly confounding factors, and (2) the difference in strength of inference between two quasi-experimental designs (Table 2, 3b and c; inference levels 3 and 4).

3. A Comparative Observational Study: Static Group Comparisons with Multiple Observations Plus Independent Evidence: Ovenbird

Cases in which there are no clear environmental perturbations and no data sets for underlying environmental variation among strata call for the examination of population trends among populations to identify patterns and then the development of hypotheses about causes. For example, in Breeding Bird Survey data for 20 physiographic strata for the Ovenbird, the overall 25-year trends are estimated to be substantial (an average change of more than one bird per route) in nine strata (Table 3). Of these nine strata, the Blue Ridge Mountains and the Adirondack Mountains are both highland areas. The design of comparisons between highland and lowland strata is basically a static group comparison with multiple observations where X is highland areas,

X	0	0	0	0
	0	0	0	0,

by itself an inadequate design (Table 2, 1c). However, in a recent multispecies analysis using BBS data for the Ovenbird and 19 other species of wood warblers, we showed that highland strata in the eastern and central United States tend to have more declining species generally than do other physiographic strata (James et al. ms). This result is too weak for causal inference, but it suggests that research is



warranted to determine what correlates of altitude might be affecting populations of warblers. The hypothesized treatment (X) is something that is present in highland areas that is not present in lowland areas. The next step should be analyses for each species that explore different combinations of levels of hypothesized causes. For example, if fragmentation, cowbird parasitism, or atmospheric pollution are the most likely suspected causes, one should construct data sets for different levels of these factors among strata, and possibly changes in levels within the BBS period, and then strengthen the design from a static group comparison by treating it as a multiple-time-series design in which different groups have different combinations and levels of these potential causes.

4. A Comparative Observational Study, Multiple Time Series With Different Levels of a Natural Intervention: Carolina Wren

Declines of populations of the nonmigratory Carolina Wren after severe winter weather is well illustrated in Breeding Bird Survey (BBS) data. In many places changes in the size of the breeding population seem to be primarily a function of the severity of the previous winter (Robbins et al. 1986, Fig. 36). The particularly severe snow and ice storms in the winters of 1976-77 and 1977-78 were associated with sharp declines as far south as the Coastal Plain (Fig. 3a) and were most dramatic in three physiographic strata, the Ohio Hills, Lexington Plain, and Highland Rim (Fig. 3b). If BBS data are interpreted as a comparative observational study, the natural intervention, with its different levels of severity in different physiographic strata, allows the researcher to compare nonlinear population trends among strata and

to make inferences about the relative effects of the two harsh winters (combined in this analysis) in strata that were affected to differing degrees. Apparently, populations in the strata most affected did not fully recover until the 1990's. This example has multiple naturally selected treatment and control groups, all with a set of pretest and posttest observations. As organized here, each group (row) is the average number of Carolina Wrens per BBS route in a stratum across years. The design has groups that received different levels of the same treatment (X), the two severe winters, which varied in intensity in different strata

0	0	0	0	$X_0$	0	0	0	0
0	0	0	0	$X_1$	0	0	0	0
0	0	0	0	$X_2$	0	0	0	0
0	0	0	0	$X_3$	0	0	0	0

#### CAUSAL INFERENCES ABOUT POPULATION TRENDS IN NEOTROPICAL MIGRANTS

What are the implications of the previous discussion for the conservation of Neotropical migrant birds? Concern for Neotropical migrants increased in the 1980's with several reports of declines in suburban forest study plots in the eastern United States (Terborgh 1989). Vague correlations with tropical deforestation were interpreted as causal by some authors, correlations with forest fragmentation on the breeding grounds by others (see review by Askins et al. 1990). Brittingham and Temple (1983) inferred a causal basis for a purported negative correlation between the abundance of the brood-parasitic Brown-headed Cowbird and the abundances of Neotropical migrants. Askins and Philbrick (1987) incorporated alternative causal factors into a

multiple-regression analysis of census data for forest-interior, long-distance migrants in a preserve in Connecticut and concluded that the degree of habitat isolation of the study area was the main factor involved in regulating their populations. Their design was a one-group pretest-posttest. On the basis of field experiments, Wilcove (1985) found that predation on artificial nests was higher in edge than in forest-interior habitats and inferred that general population declines were due to predation associated with forest fragmentation.

Most researchers acknowledge that the best source of data for the documentation of regional population trends in North American land birds is the Breeding Bird Survey, the roadside census program started in 1966 and sponsored by the U.S. Fish and Wildlife Service (FWS) and the Canadian Wildlife Service (Robbins et al. 1986). The most recent FWS analyses (Robbins et al. 1989; Peterjohn et al. this volume) of Breeding Bird Survey data suggest that a major feature of the data is that populations of Neotropical migrants increased in the 1970's and decreased in the 1980's. This phenomenon has received no attention in terms of causal analysis. However, we have not been able to find this peak in our own analyses of BBS data for wood warblers (James et al. ms).

Robbins et al. (1989) and Bohning-Gaese et al. (1992) demonstrated associations between population trends and causes using regression-type analyses with BBS data. We have reservations about their analyses, however, because they assumed that data on trends in different species were independent. This is unlikely to be true because BBS data for all species come from the same routes run at the same times.

In all of the above literature, purported multispecies declines were attributed to a preferred alternative cause, and the design of the research did not adequately rule out alternative explanations. In addition to their weak designs, the validity of extrapolations from population trends detected in local studies to scales larger than their immediate surroundings have rarely been checked. For example, there is no evidence in Breeding Bird Survey data that cowbirds have been increasing in abundance in the eastern United States in the last 25 years (John Sauer, David Wiedenfeld, personal communications) or even that Neotropical migrant forest birds as a group are declining in New England (Smith et al. 1993). Another relevant issue is whether observed variation in survival (for example, due to predation) actually affects the size of the population in the following breeding season (Newton 1980). Unfortunately, we do not have complete information about the causes of population limitation for even one species.

At least three major problems must be addressed before valid inferences can be developed about what factors affect population trends of Neotropical migrants. The first, of course, is accurate documentation of their fluctuations. This is fairly easily accomplished at a local scale, but determination of regional population trends requires sampling at that scale. The second problem, which is the one discussed in this paper, is how to determine the major causes of observed population trends. If experiments are feasible, they may show what environmental factors, if modified, would cause a change in the number of individuals in the next breeding season in a local population. Even so, their results can be misleading. The validity of causal

inferences derived from experiments depends upon their design. In addition, extrapolation of the results of local field experiments to large geographic areas may not be justified. We think that carefully designed comparative observational studies are called for. The third problem is that confirmed sources of mortality are not necessarily causes of population regulation. There is a clear need for well-designed experiments that allow for the identification of processes that, when modified, result in changes in the sizes of the breeding populations.

#### RECOMMENDATIONS

1. Study individual species that have different population trends in different areas. The geographic variation in population trends within species can be used to advantage in causal analyses. Construct data sets for different levels of potential causes in the different areas. Seek examples in which levels of treatments change during a time series. Analyze the data for trends and causes as comparative observational studies that use the strongest possible levels of inference, preferably with the multiple-time-series design. The Breeding Bird Survey is an appropriate source of data for the population trends. In statistical analyses of Breeding Bird Survey data, do not treat species as independent units.
2. If feasible, conduct field experiments to test whether purported causes actually affect the size of the breeding population.

3. Be explicit about the level and scope of causal inference allowed by the design of a study, including the extent to which alternative explanations have been accounted for.

#### CONCLUSION

At present, ecological research intended to discover the causes of broad-scale ecological phenomena, such as long-term population trends in birds, is poorly developed, and there is a large gap in knowledge at the research-management interface. Until sound inferences can be developed about what environmental factors, when modified, will make a difference, conservation and management will be inefficient. An important part of the development of such causal inferences should be the incorporation of principles of quasi-experimental design into broad-scale observational studies. This type of research deserves more attention.

## ACKNOWLEDGMENTS

We thank James Cox, Duncan Evered, Charles Hess, Don Levitan, Arie van Noordwijk, Noel Wamer, David Wiedenfeld, and two anonymous reviewers for their helpful criticism of earlier drafts of the manuscript.

## LITERATURE CITED

- Askins, R. A., J. F. Lynch, and R. Greenberg. 1990. Population declines in migratory birds in eastern North America. Pages 1-57 in D. M. Power, editor. Current ornithology, Vol. 7. Plenum Press, New York.
- Askins, R. A., and M. J. Philbrick. 1987. Effects of changes in regional forest abundance on the decline and recovery of a forest bird community. Wilson Bull. 99:7-21.
- Bohning-Gaese, K., M. L. Taper, and J. H. Brown. 1992. Are declines in North American insectivorous songbirds due to causes on the breeding range? Conserv. Biol. 7:6-86.
- Brittingham, M. C., and S. A. Temple. 1983. Have cowbirds caused forest songbirds to decline? BioScience 33:31-35.
- Campbell, D. T., and J. C. Stanley. 1966. Experimental and quasi-experimental designs for research. Rand McNally, Chicago.
- Cochran, W. G. 1983. Planning and analysis of observational studies. Wiley, New York.
- Cook, T. D., and D. T. Campbell. 1979. Quasi-Experimentation: design and analysis issues for field settings. Rand McNally, Chicago.
- Diamond, J. 1986. Overview: laboratory experiments, field experiments, and natural experiments. Pages 3-22 in J. Diamond and T. J. Case, editors. Community ecology. Harper and Row, New York.
- Eberhardt, L. L., and J. M. Thomas. 1991. Designing environmental field studies. Ecol. Monogr. 61:53-73.
- Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, New York.



- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54:187-211.
- James, F. C., D. A. Wiedenfeld, and C. E. McCulloch. Manuscript. Analysis of population trends in warblers using two nonlinear methods (submitted to *Ecol. Appl.*).
- Kish, L. 1987. Statistical design for research. Wiley, New York.
- Maclure, M. 1988. Refutation in epidemiology: why else not? Pages 130-138 in K. J. Rothman, editor. Causal inference. Epidemiology Resources, Chestnut Hill, Massachusetts.
- Newton, I. 1980. The role of food in limiting bird numbers. *Ardea* 68:11-30.
- Peterjohn B. G., J. R. Sauer, and C. S. Robbins (this volume). The North American Breeding Bird Survey and population trends of Neotropical migrant birds.
- Probst, J. R. 1986. A review of factors limiting the Kirtland's Warbler on its breeding grounds. *Am. Midl. Nat.* 116:87-100.
- Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. The Breeding Bird Survey: its first fifteen years, 1965-1977. U. S. Fish and Wildlife Service, Resource Publ. 157, Washington, D.C.
- Robbins, C. S., J. R. Sauer, R. S. Greenberg, and S. Droege. 1989. Population declines in North American birds that migrate to the Neotropics. *Proc. Nat. Acad. Sci. USA* 86:7654-7662.
- Rothman, K. J. 1988. Causal inference. Epidemiology Resources, Chestnut Hill, Massachusetts.
- Scheiner, S. M., and J. Gurevitch, editors. 1993. Design and analysis of ecological experiments. Chapman and Hall, New York.

- Skalski, J. R., and D. S. Robson. 1992. Techniques for wildlife investigations: design and analysis of capture data. Academic Press, San Diego, California.
- Smith, C. R., D. M. Pence, and R. J. O'Connor. 1993. Status of Neotropical migratory birds in the Northeast: a preliminary assessment. Pages 172-188 in D. M. Finch and P. W. Stangel, editors. Status and management of Neotropical migratory birds. USDA Rocky Mountain Forest and Range Experiment Station, Ft. Collins, Colorado.
- Terborgh, J. 1989. Where have all the birds gone? Essays in the biology and conservation of birds that migrate to the American tropics. Princeton University Press, Princeton, New Jersey.
- Tinbergen J. M., J. H. Van Balen, and H. M. Van Eck. 1985. Density dependent survival in an isolated Great Tit population: Kluyver's data reanalyzed. *Ardea* 73:3-48.
- Walters, J. R., C. K. Copeyon, and J. H. Carter, III. 1990. Test of the ecological basis of cooperative breeding in Red-cockaded Woodpeckers. *The Auk* 109:90-97.
- Wilcove, D. S. 1985. Nest predation in forest tracts and the decline of migratory songbirds. *Ecology* 66:1211-1214.

Table 1. Definitions, based largely on Campbell and Stanley (1966), Cook and Campbell (1979), and Cochran (1983).

---

comparative observational study -- a study involving sampling of naturally occurring events but employing elements of quasiexperimental design such as control groups to gain inferences about causal relationships; the same as the "ex post facto experiments" of Campbell and Stanley (1966) and "analytical sampling" by Eberhardt and Thomas (1991); also includes the "natural experiments" of Diamond (1986) in those cases in which rival hypotheses can be rendered implausible

control group -- an experimental unit that did not receive the treatment; in a comparative observational study, a group to be compared with another group that was subjected to a natural intervention

controlled experiment -- a study in which the researcher applies a treatment, as opposed to an observational study or a survey

descriptive survey -- a sampling study not planned for the analysis of causes; nevertheless the results could be analyzed using quasiexperimental designs, especially if the effects of distinct perturbations are evident; the Breeding Bird Survey is an example

experimental design -- an arrangement of observational units and treatments planned to detect the effect of a treatment or perturbation

Table 1. (continued)

---

intervention -- an event that occurred within a series of observations of the same group; may be a treatment applied by a researcher or a naturally occurring event

multiple-time-series design -- a comparative design involving more than one group, each of which has several observations before and after a treatment

observational study -- a study consisting of sampling; the study is uncontrolled in the sense that the researcher did not apply a treatment (Cochran 1983); it can be simply a descriptive survey without any intent toward causal inference or a comparative observational study planned to analyze causes; it can encompass the entire domain of interest more easily than can most experiments

quasiexperiment -- an experiment in which there is no randomization of the assignment of experimental units to treatments (Cook and Campbell 1979); same as the "pseudoexperiments" of Cochran (1983)

randomization -- in experimental design, the random assignment of experimental units to treatments, often not possible in field experiments

replication -- repeated cases of the application of treatments to experimental units, whether or not they are under the control of the observer

time-series design with intervention -- a design consisting of several observations before and after an intervening treatment

Table 1. (continued)

---

true experiment--an experiment whose design includes control groups,  
randomization and replication; for an example of a true field  
experiment with birds, see Walters et al. (1990)

---

Table 2. Examples of four categories of research designs used for causal analysis (from Campbell and Stanley 1966).

Design	Notation <sup>1</sup>	Strength of				
		Treatment controlled by observer	Randomization	Control group	causal inference (0-5) <sup>2</sup>	Domain (1-3) <sup>3</sup>
1. Inadequate designs						
a. One-shot case study	X 0	+	-	-	0	1-2
b. One-group pretest-posttest	0 X 0	+	-	-	1	1-2
c. Static-group comparison	X 0					
	0	+	-	+	1	1-2
2. A true experimental design	R 0 X 0					
	R 0 0	+	+	+	5	1
3. Quasiexperimental designs (nonrandomized control)						
a. Nonequivalent-control-group design (removes effect of time)	0 X 0	+	-	+	3	2
	0 0					

Table 2 (continued).

Design	Notation <sup>1</sup>	Treatment		Control group	Strength of causal inference		Domain (1-3) <sup>3</sup>
		controlled by observer	Randomization		(0-5) <sup>2</sup>		
b. Single time series with intervention	0000 X 0000	+	-	-	3		2
c. Multiple time series with intervention	0000 X 0000						
	0000 0000	+	-	+	4		2
4. Comparative observational studies <sup>4</sup>							
a. Nonequivalent-control-group design	0 X 0	-	-	+	1		2-3
b. Single time series with natural intervention	0000 X 0000	-	-	-	2		2-3

Table 2 (continued).

Design	Notation <sup>1</sup>	Treatment		Control Group	Strength of causal inference		Domain (1-5)
		controlled by observer	Randomization		(0-5)		
c. Multiple time series with natural intervention	0000 X 0000 0000 0000	-	-	+	3	2-3	

<sup>1</sup>X = treatment; 0 = observation; R = units randomly assigned to treatments or controls; rows are for single groups; columns are for simultaneous events.

<sup>2</sup>Categories, assigned by the authors, range from 0 for none to 5 for maximal causal inference.

<sup>3</sup>Categories range from 1 for the usual restricted geographical (or temporal) domain of a field experiment to 3 for the entire domain of interest.

<sup>4</sup>Same designs as in category 3, quasiexperimental designs, but X's are naturally occurring events.



Table 3. Population trends in the Ovenbird from 1970-72 to 1986-88 by 20 physiographic strata, as analyzed by semiparametric route regression limited to 300 iterations.<sup>1</sup> All strata are reported that had differences of an average of more than one bird per route. An asterisk means statistical significance at  $\alpha =$  less than 0.05.

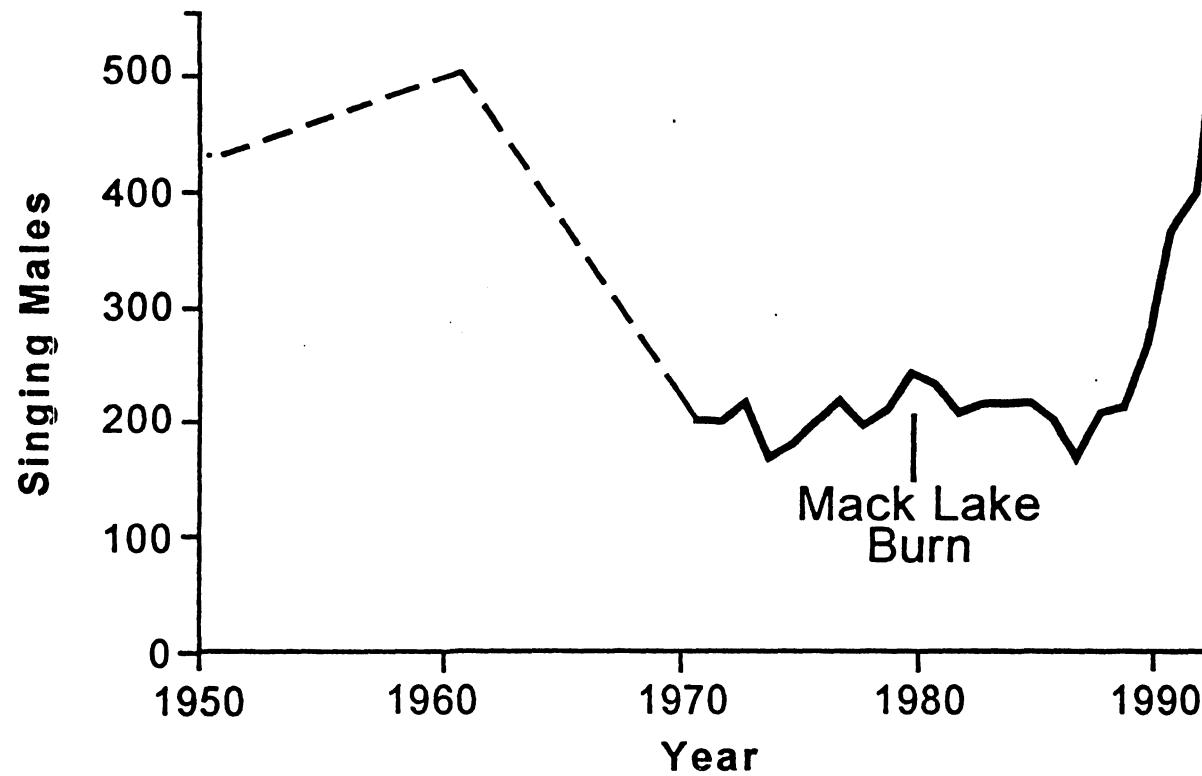
Physiographic stratum	df	Average change in birds per route
Blue Ridge Mountains	2	-5.6*
Adirondack Mountains	4	-2.8
S. Upper Coastal Plain	28	1.1
Spruce Hardwoods and Boreal Forest	67	1.2
Great Lakes Transition	19	1.5
St. Lawrence River Plain	19	2.1*
S. New England and Glaciated Coastal Plain	23	2.3
Allegheny Plateau	36	2.4*
Northern New England	20	3.6

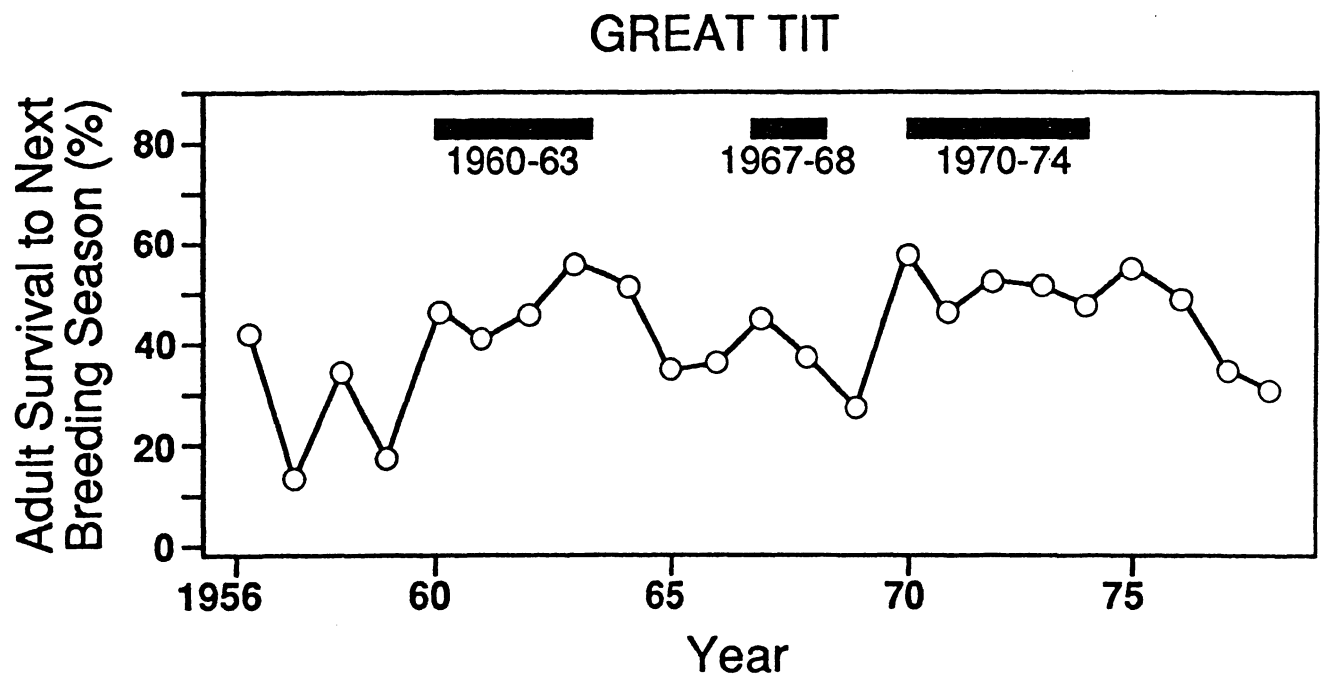
<sup>1</sup>Methods of data analysis are described by James et al. (ms).

## FIGURE LEGENDS

- Fig. 1. The number of singing male Kirtland's Warblers (Dendroica kirtlandii) stabilized after the Brown-headed Cowbird (Molothrus ater) control program began in 1970 and rose sharply after 1990, when ecological succession following the Mack Lake Burn provided new suitable habitat.
- Fig. 2. The rate of adult survival to the next breeding season in a population of Great Tits (Parus major) using nest boxes on the island of Vlieland, The Netherlands. The black bars show years when juvenile recruitment was reduced by removal of eggs or birds. Based on Tinbergen et al. (1985).
- Fig. 3. Population trends in the Carolina Wren (Thryothorus ludovicianus) in ten physiographic strata as estimated by nonlinear semiparametric route regression (James et al. ms) applied to data from the Breeding Bird Survey, but without smoothing.

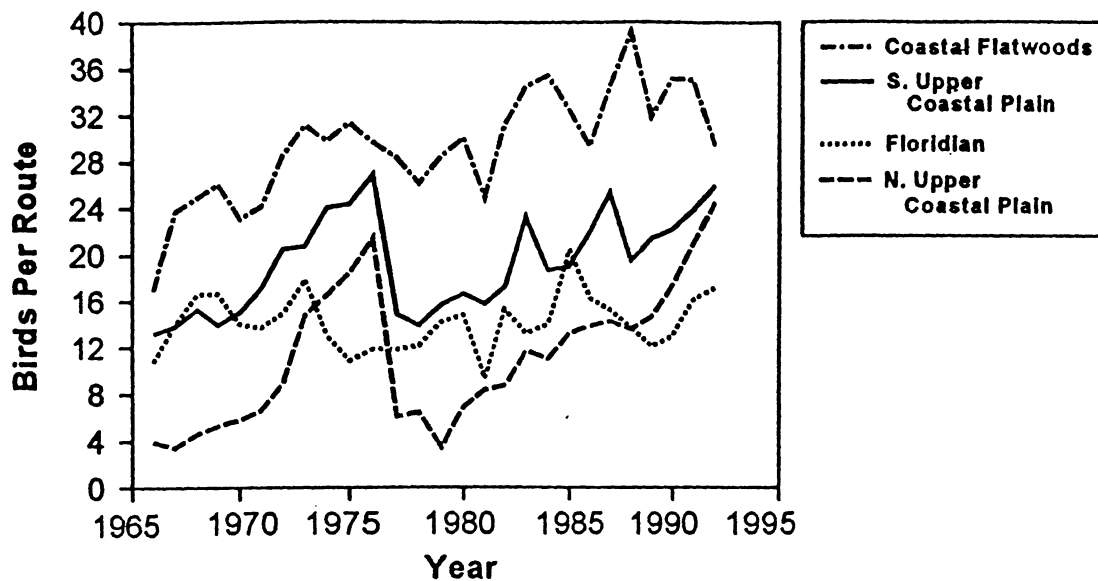
# KIRTLAND'S WARBLER CENSUS





# CAROLINA WREN

a.



b.

